
Automation vs Interaction: What's Best for Big Data? Both Are Not Enough!

Samuel P. Uselton

Center for Applied Scientific Computing

October 29, 1999



What is the REAL question?

- The data doesn't *care!* What is best **FOR USERS?**
- Interactive Exploration
 - Better Understanding
 - Active Driving - Engagement
 - True Exploration - Finding and Probing the Unexpected
- Automatic Analysis
 - User Efficient
 - Focus on Relevant Data
 - Offline Preparation
 - Effective Use of Computational Resources

Users Need BOTH And MORE!

- **Usual “Combined” Strategy**
 - **Complementary Role Players...**
.... Like Fresh Horses for the Pony Express
- **Better Combined Strategy**
 - **Both IN TANDEM**
.... Like Clydesdales pulling the giant ASCI load

Estimating Visualization Bandwidth

- **Display**

- 2K x 2K x 24 bits
- Corresponds to 3 scalar fields, 8 bits each

- **Frame Rate**

- 10 frames per second

Result:

120 Megabytes per second
(a generous OVERestimate)

Data Production Capability: sPPM Example

- Timestep produced every 30 seconds
- Eight Billion (!) cell mesh
- Five Floating Point Values per Cell

Result: > Five BILLION bytes per Second

Conclusion: Data Production Is
Forty(40) TIMES
Visualization Bandwidth

Limits: ...and the human will get tired first!

- **Remembering and Correlating**
- **Tired - Bored - Saturated**
- **Environmental Distractions**

Human parallelism? Obstacles:

Communication and Coordination

Correlation Across Team

Individual Differences

**Interaction: Longer Looks at the Same Data
.....EVEN MORE Data Never Seen**

Another Limit: **Storage** (sPPM Example)

- **Save every kth time step**
 - 10,000 timesteps computed; 300 saved
- **Save at a subset of grid cells**
- **Save only some variables**
 - one derived scalar, not the original five
- **Save variables at lower precision**
 - 8 or 16 bit integer, not floating point
- **Resample to lower resolution mesh**
- **AND the percentage stored is likely to decline**

Automatic Analysis

- **Features - Being smarter about what to STORE**
- **Completely automatic ?**
 - "Movies"
 - Concise summaries - statistics, simple plots, ...
- **Powerful when appropriate, but ...**

We ask each other "How did you get this answer?"

We need the ability to ask the same question of automated tools.

A Place to Start: WHY is the Data SO BIG?!

- **To Answer Questions:
A Larger DATA ANALYSIS Context**
- **Interactive or Automatic Better FOR WHAT?**
- **Data Use: "Scientific versus Engineering"**

BEST is to use them BOTH, and TOGETHER!

- **Automatic Identification of Regions WORTH Interactive Exploration**
- **Automatic Identification of Regions worth SAVING in DETAIL**
- **Migrate Techniques From Interactive to Automatic**

BE CREATIVE!

Acknowledgements

- **Current Colleagues**
 - CASC, LLNL, ASCI, Tri-Lab, DoE, ...
- **Former Colleagues**
 - NASA Ames DAGroup, and others
- **Other Stimulating Conversationalists**